Accelerometer-based real-time voice activity detection using neck surface vibration measurement

A THESIS
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Saurav Kumar Dubey

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Dr. Jürgen Konczak

June 2019

Acknowledgment

Dr. Jürgen Konczak

Arash Mahnan

Jinseok Oh

Jiapeng Xu

Jessica Holst-Wolf

Naveen Elangovan

All members of the Human Sensorimotor Control Laboratory

Abstract

Speech analysis has a growing number of clinical and industry applications, all of which rely on Voice Activity Detection (VAD). Common VAD applications use microphones, which can be problematic in the presence of background noise and additional voices. Recent studies have utilized accelerometers instead of microphones as voice transducers. As part of a larger research project on impaired speech in the voce disorder spasmodic dysphonia (SD), this study aimed to explore the use of wearable accelerometers to detect speech. These accelerometers would be part of a real-time VAD system embedded in a wearable neck collar for patients with SD. This collar would deliver vibro-tactile stimulation (VTS) to the laryngeal muscles during speech as a therapy for these patients. The aims of this research concerned a) finding the ideal location on the neck to place the accelerometers and b) developing a VAD algorithm that reliably detects the onset and offset of speech based on these accelerometer signals.

Methods: 6 healthy adult participants (M/F = 3/3, 26 ± SD = 5.1 years) vocalized 20 sample sentences under 12 conditions from a combination of 3 variables: 1) Normal or slow speed of speech, 2) Three accelerometer attachment locations: thyroid cartilage, sternocleidomastoid, and superior to the C7 vertebra, and 3) Application of VTS during speech in two locations. Time-synchronized acceleration and audio were recorded in each condition.

Results: Number of onsets of voice activity and total time voiced, as calculated from application of the VAD algorithm to the acceleration data, were measured. The thyroid cartilage attachment location had over 90% accuracy detecting speech in both measures

on average. The average accuracy of the sternocleidomastoid location was below 75% accuracy and was below 15% for C7.

Discussion: Placing of an accelerometer at the thyroid cartilage for real-time detection of speech was shown to be feasible. The obtained usability data document that accelerometer signals at this anatomical landmark provide the most reliable data to detect speech. The other two locations tested were too variable in accuracy for implementing VAD. With respect to using the established VAD algorithm in the wearable collar device to treat voice symptoms in spasmodic dysphonia, one needs to state that the algorithm can be improved in robustness to filter out the noise caused by vibration. The use of advanced processing methods such as adaptive filtering will likely deliver the desired result.

Table of Contents

List of Tables

List of Figures

Other (list of abv.)

Voice Activity Detection – VAD

Spasmodic Dysphonia – SD

Vibro-tactile Stimulation – VTS

Ambulatory Phonation Monitor – APM

Energy – E

Average Power – P

Spectral Energy Density – $E_s(f)$

Power Spectral Density – P_{xx}

Average Spectral Power – P_f

Introduction

Voice Activity Detection

Speech analysis has a variety of applications ranging from software for automatic speech recognition in industry to diagnosing voice and speech disorders in clinical settings. All of these uses rely on Voice Activity Detection (VAD) to differentiate utterances from silence during speech. Until recently, microphones have been widely used as voice transducers in VAD apparatuses, but they can be difficult to implement outside of a controlled laboratory or clinical environment due to background noise. For example, exposure to other voices or ambient sound from the environment in the field can require extensive signal processing for offline speech analysis and make real-time VAD difficult. Actively filtering this noise from a microphone signal in real-time can be complicated and computationally inefficient. Interest in accelerometers to collect voice activity data, especially in noisy environments, has increased to combat the problems with using microphones. In comparison to binaural microphone recording, a commonly accepted method to reduce recorded background noise, a neck-attached accelerometer captured voice signals significantly more accurately in low signal-to-noise ratio environments (Lindstrom, Ren, Li, & Waye, 2009). Even in the presence of 67.5 decibel background noise, compared to the 40 decibel volume tested in the previously mentioned study, a neck-attached accelerometer was resistant to noise while recording laryngeal vibrations (Yiu & Yip, 2016). Both of these studies support accelerometers' robustness to environmental noise.

Research has demonstrated the feasibility of using accelerometers to monitor voice activity reliably and validly. Neck surface vibration magnitude correlates with sound pressure level (Švec, Titze, & Popolo, 2005) and fundamental frequency during speech (Askenfelt, Gauffin, Sundberg, & Kitzing, 1980; Szabo, Hammarberg, Hakansson, & Sodersten, 2001). Both of these parameters are easily calculated and correlated with speech activity, validating the accelerometer for use in determining vocalization. Supporting the reliability of acceleration signals in VAD, an accelerometer was used to detect the onset of speech, after a trigger event, in pseudo-real-time with 96% accuracy (Vitikainen, Makela, Lioumis, Jousmaki, & Makela, 2015). This was not truly real-time as a trigger event allowed for a predictable window in which to expect speech onset. However, this study demonstrates that neck surface vibrations as measured by accelerometers can plausibly be used as reliable markers for speech activity. Outside of these studies that were conducted in a controlled laboratory setting, accelerometers have been used in workplace environments with random background noise to track speech activity throughout a day (Matic, Osmani, & Mayora, 2012). In these studies, accelerometers have recorded speech activity as well as or better than microphones in controlled and field environments with constant or random noise.

Because accelerometers have been utilized for VAD, their implementation has evolved drastically. Initial models like the Ambulatory Phonation Monitor (APM - KayPENTAX, Lincoln Park, NY, United States) use a separate hardware module for collecting and storing accelerometer data. This process has evolved to the ability to use a smartphone as the data acquisition device (Mehta, Zanartu, Feng, Cheyne II, & Hillman, 2012). In any case, the main clinical purpose of monitoring voice activity has been the measurement of phonation time, or vocal dosimetry, in healthy individuals. These accelerometer-based vocal dosimetry studies allow for detection of speech characteristics related to vocal disorders (Mehta et al., 2012), or phonation time as an indicator of social wellbeing (Matic et al., 2012).

While the majority of vocal dose research focuses on populations with healthy speech production, the devices have been tested for use with patients of spasmodic dysphonia (SD), a disorder where producing speech is difficult due to involuntary spasms of the laryngeal muscles. Issues with microphone recordings of healthy voices are exacerbated in dysphonic patients due to the strained speech produced, often at a much lower volume than is normal for the healthy population. As dysphonia severity increases, the effectiveness of microphone-based measures decrease (Hillman, Heaton, Masaki, Zeitels, & Cheyne, 2006). Accelerometers, however, are able to capture acceleration data from speech in even the most severe cases of dysphonia (Cheyne, Hanson, Genereux, Stevens, & Hillman, 2003). For this reason, accelerometer-based voice dosimeters have been proposed as tools not only for voice activity assessment to detect speech disorder precursors, but also for potentially providing real-time biofeedback to dysphonic patients based on their speech activity (Mehta et al., 2013; Nacci et al., 2013).

Signal Processing

Detecting voice activity from acceleration data will require signal processing to filter the data and set thresholds for speech. There are numerous signal processing methods proposed for VAD in current research, varying in complexity and efficiency. A basic method of analyzing speech signals for VAD is calculating the average power of a signal in the time or frequency domain. A signal can be characterized in terms of energy and power. In the time-domain, the energy, E, of a signal $x(n)$ is the sum of each squared value in the signal:

$$E = \sum_{n=-\infty}^{\infty} |x(n)|^2 \tag{1}$$

The average power, P, of this signal, then, is the energy of the signal over time. In a discrete-time signal, this can be represented by the number of samples, N, over which the average power is calculated:

$$P = \frac{1}{N} \sum_{n=0}^{N-1} |x(n)|^2 \tag{2}$$

This calculation can also be conducted similarly in the frequency domain to find spectral energy density, power spectral density, and average spectral power. The spectral energy density, $E_s(f)$ can be calculated using the Fourier transform, $X(f)$, to represent the original signal in the frequency domain as a relationship between frequencies and their magnitudes present in the signal:

$$E_s(f) = |X(f)|^2 \tag{3}$$

The power spectral density of the signal, P_{xx}, describes how the power of a signal is distributed across all frequencies in the signal, where Δt represents the sampling interval:

$$P_{xx}(f) = \frac{\Delta t}{N}\left|\sum_{n=0}^{N-1} X(f)_n\right|^2 \tag{4}$$

The average spectral power of this signal in the frequency domain, P_f, can then be calculated by integrating the power spectral density over all frequencies in the signal, where f_s is the sampling frequency:

$$P_f = \int_{-f_s/2}^{f_s/2} P_{xx}(f)\, df \tag{5}$$

According to Parseval's Theorem, the energy of the signal in the time domain is equal to the summation of all frequency components of the spectral energy density of the signal (Schafer & Oppenheim, 2010). Thus, the calculated average power of a window of samples in the time and frequency domain would be equal as well. When analyzing healthy voices in a laboratory setting, then, the more efficient amplitude based calculations of average power would be ideal in differentiating voiced from unvoiced windows of a signal. However, applying this VAD system to the real world could introduce unwanted noise, such as movement artifacts. Using the frequency domain allows for simple filtering of these extra frequencies.

When using either measure for VAD, a threshold level is set below that of phonation to register any speech activity. Depending on the application of this VAD system, these thresholds can be generalized to function with any individual, or each threshold can be individualized based on a calibration process. For individual calibration, a baseline value can be measured by collecting data without speech activity and calculating a threshold metric, such as average signal power.

This study proposes the use of an accelerometer to measure neck surface vibrations and detect the onset of speech in real-time. This study is part of a larger project with the goal of creating a wearable collar that delivers vibration therapy to SD patients as they speak (Mahnan, Faraji, & Konczak, 2019). A persistent improvement in voice quality has been demonstrated in SD patients following the application of vibro-tactile stimulation (VTS) to the laryngeal muscles (Khosravani et al., 2019). While existing devices are able to utilize either amplitude or frequency based algorithms for VAD, none have done so with concurrent vibration in the vicinity of the accelerometer as this study proposes. The introduction of this vibration adds noise to the accelerometer's recording, which could interfere with the existing VAD techniques. The signal processing workflow in this study will need to filter out the laryngeal muscle vibration while implementing a VAD algorithm on the filtered data.

Specific Aims

Aim 1: The first aim of this study is to determine the accelerometer attachment location providing the best neck surface vibration signal amidst concurrent vibration at the laryngeal muscles.

This location must be within the region of a neck collar circumscribing the neck at the thyroid cartilage. The vocal dosimeter devices discussed thus far use an accelerometer attachment at the jugular notch to obtain the strongest acceleration signal from neck surface vibrations during speech. While this location does not fall within the region of the desired collar, certain areas in this region do provide strong magnitudes of acceleration normal to the skin's surface (Nolan, Madden, & Burke, 2009).

Aim 2: The second aim of this study is to develop a signal processing algorithm with the capability of processing the acceleration data to detect speech in real-time alongside the delivery of VTS during speech.

Available voice dosimeters, with the exception of the APM, do not have real-time signal processing capabilities. While the APM can detect speech based on real-time vocal characteristics, it is based on loudness measured by signal amplitude, not the onset of speech itself (Nacci et al., 2013). Accelerometer-based VAD algorithms that do exist are not designed for use with concurrent vibration, which could cause false detections of speech activity.

Methods

Participants

Voice and acceleration data were collected from 6 healthy adult participants (M/F = 3/3, 26 ± 5.1 years) in the study. Healthy was defined as lacking any self-identified neurological, movement, or speech disorders using a Subject Information Form (see SUPPLEMENTARY MATERIALS). Participants were recruited using flyers posted around the University of Minnesota campus. The experimental protocol was approved by the University of Minnesota Institutional Review Board (STUDY00004785). All participants gave their informed consent before partaking in the study. The study was conducted at the Human Sensorimotor Control Lab in one session per participant.

Instrumentation

Audio data were recorded at 44100 Hz using an ECM-88B Electret Condenser Microphone (Sony Corporation, Tokyo, Japan) connected to a MixPre-6 microphone preamplifier. The preamplifier was connected to a computer using Audacity (Audacity Team, 2019) to record the signal. The accelerometer used was the BU-27135-000, a single-axis accelerometer (Knowles Electronics LLC, Itasca, IL, United States). The acceleration data was collected and recorded directly to an SD card by an Arduino Uno at 1000 Hz. This sampling frequency accounts for the average fundamental frequency of vocalized vowels in males and females of up to 400 Hz (Stevens, 2000). The accelerometer was connected to the Arduino Uno using three insulated multiple strand wires. The accelerometer output was a voltage value as an indication of acceleration. Note that the accelerometer did not return true acceleration values, but a change in voltage caused by alteration of device's acceleration. This analog voltage signal was converted by the Arduino to an integer value on a scale from 0 to 1023, corresponding to a recorded voltage of 0 to 5 Volts divided by a reference voltage of 1.1 Volts. The resulting unitless acceleration signal was then normalized by subtracting the mean of the entire signal to align the value to an amplitude of 0.

Acceleration and audio recordings were time-synchronized in order to validate VAD from the accelerometer data. Time synchronization was implemented by producing a 1000 Hz beep for 250 milliseconds at the start and end of accelerometer signal recording that was recorded in the audio signal. The audio signal was then trimmed to the start of the first and last beep. Laryngeal vibration was provided by 2 Pico Vibe™ 9 millimeter vibration motors (Precision MicrodrivesTM, Model 307 – 100). The motors operated at a stimulation frequency of 100 Hz, supplied with 1.1 volts from an adjustable voltage power supply. The accelerometer and vibrators were attached to the skin of the neck using double sided tape and one piece of Blenderm tape (3M, Maplewood, MN, United States) on top of each device to secure them to the skin (see Figure 1).

Experimental Procedure

The accelerometer was attached to three locations: thyroid cartilage below the thyroid notch, sternocleidomastoid in line with the thyroid notch, and 2.5 centimeters above C7 on the back of the neck. Vibratory motors were attached bilaterally over the laryngeal area, lateral to the thyroid cartilage (see Figure 1A).

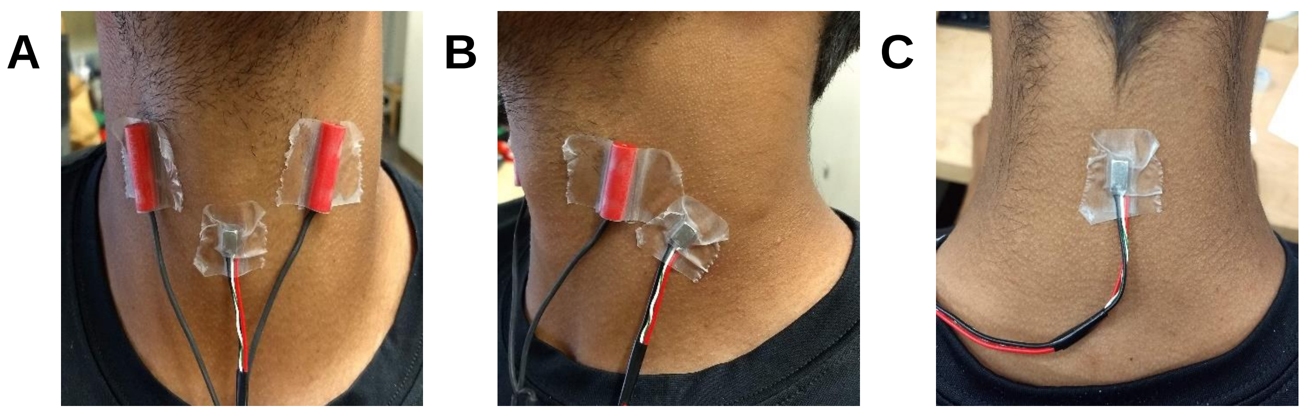

Figure 1. Attachment of vibratory motors to the laryngeal muscles lateral to the thyroid cartilage and accelerometer attachment locations: (A) thyroid cartilage, (B) sternocleidomastoid lateral to the thyroid cartilage, and (C) 2.5 centimeters above C7.

For the thyroid cartilage, the accelerometer was attached on the center of the thyroid cartilage, below the hyoid bone. For the sternocleidomastoid, participants were asked to rotate their head ninety degrees and the accelerometer was placed lateral to the protrusion of the sternocleidomastoid and in line with the thyroid cartilage. For the C7 location, participants were asked to look down and the accelerometer was placed 2.5

centimeters superior to the protrusion of the C7 vertebra (see Figure 1). The orientation of the accelerometer was irrelevant because it only recorded acceleration in one axis, perpendicular to the surface of the skin.

The experimental protocol consisted of speaking 20 sample sentences (see SUPPLEMENTARY MATERIALS) in 12 different conditions of 3 variables: accelerometer attachment location, speed of speech, and application of VTS (see Figure 2). Speech was tested at normal and slow speeds to account for variation in individual speaking styles when developing the VAD algorithm. Slow speech was designated as pauses between words or phrases in each sentence (e.g. “Tom – wanted – to be – in – the army”). Accounting for this choppiness in speech would accommodate natural pauses as well as speech in SD patients. VTS was applied while recording speech to design a VAD algorithm robust to the laryngeal vibration frequency. Speech data was collected with and without VTS application at all accelerometer attachments. A baseline trial without speech was recorded at each accelerometer location (see Figure 2).

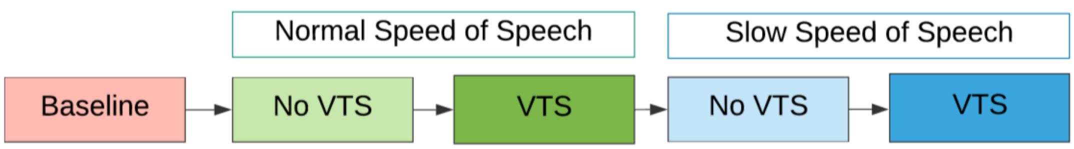

Figure 2. Experimental protocol at each accelerometer attachment location. Data were collected in this order: a baseline trial with no speech, normal speed of speech, and then slow speed of speech. For each speed of speech, data were recorded in separate trials with and without the application of VTS.

Signal Processing

Audio data were imported from Audacity to MATLAB following testing each participant. Acceleration data were converted from binary to comma-separated value files on the Arduino, then imported from the SD card to MATLAB. Data analysis and signal processing was conducted in MATLAB R2018b (The MathWorks Inc., Natick, Massachusetts, United States).

Non-vibration and vibration trials were treated with different filters. Non-vibration trials were used as evidence for the feasibility of using an accelerometer at any of the chosen regions around the neck for generally applicable speech detection. Vibration trials targeted the specific application of the delivery of VTS using a collar. Thus, filters to remove the vibration frequency and noise associated with the current to the motors were only applied in trials with vibration.

Non-vibration trials were treated with a band pass filter from 80-400 Hz to remove low frequency bands caused by movement during the trial and high frequency noise from the accelerometer. Vibration trials were treated with a band pass filter from 110-400 Hz. The increase from 80 Hz to 110 Hz on the lower stop band accounts for the vibration frequency of the motors that varied from 99 Hz to 109 Hz across participants. This slight variance of vibration frequency can be attributed to attenuation of the vibration due to anatomical differences of the neck region between participants and minor inconsistency of the voltage provided to the motors. The vibration trials were then treated with three band stop filters to remove the 2^{nd}, 3^{rd}, and 4^{th} harmonic of the 60 Hz noise caused by the current to the motors. Only FIR filters were used to avoid nonlinear phase distortions. These filters were applied using the MATLAB function *filtfilt*, which produces zero-phase distortion, however, it does introduce a constant magnitude distortion. This is accounted for later in the signal processing workflow by a scale factor to adjust the threshold magnitude. Because there is a magnitude distortion and the possibility of phase distortion when using other filters, which might influence the acceleration signal in the time domain, calculations in time domain were excluded for vibration trials.

VAD Algorithm

To differentiate speech acceleration signals from no activity, a baseline trial in each location was used to calculate a threshold (see Figure 2 baseline trial). A 5000 millisecond interval of each baseline trial was divided into 50 millisecond subintervals. The average power and average spectral power were calculated for each subinterval, then averaged across all subintervals to determine the threshold for speech activity (see Figure 3).

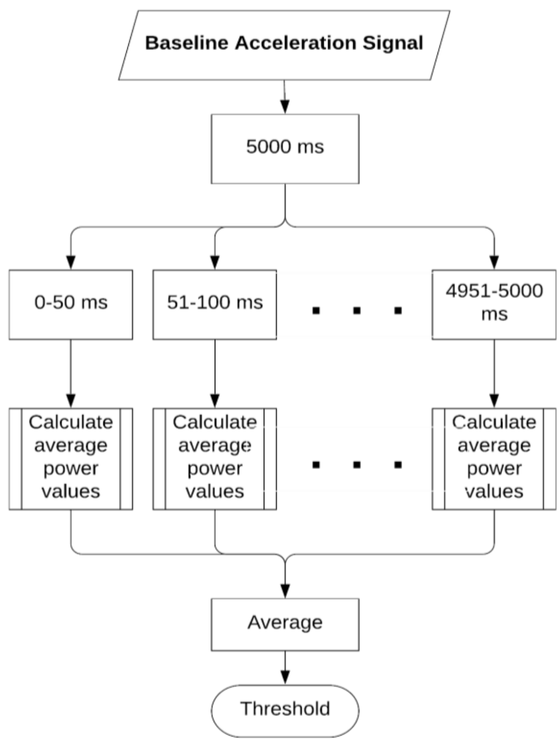

Figure 3. Threshold calculation. A 5000 millisecond interval of the baseline trial was divided into 100 subintervals of 50 milliseconds. The average power values for each of these subintervals were calculated. These average power values were then averaged to calculate the threshold for speech activity to be used by the VAD algorithm (see Figure 4)

Acceleration data were collected in speech trials following the baseline trial (see Figure 2). The VAD algorithm analyzed the acceleration signals in 250 millisecond intervals that were filtered independently to resemble how the algorithm would intake data in real-time. After filtering, the algorithm (see Figure 4) was applied to the acceleration signal intervals. The intervals were divided into 50 millisecond subintervals and the average power of each subinterval was calculated in the time and frequency domains. These two values were then compared to the threshold, and if any two contiguous subintervals were greater than the threshold, the entire interval was considered voiced. After each interval was filtered and analyzed, the complete filtered signal was returned. The length of intervals and subintervals is modular in the VAD algorithm.

Since the algorithm is designed for real-time implementation, as the entire interval needs to be analyzed before determining if it is voiced, the length of the interval is defined as the expected real-time delay in vibration onset. Longer interval lengths could result in more lenient voice activity detection, but the increased delay would become noticeable. The increased length could also cause overlap between different instances of speech, which would mark periods between speech activity as voiced. Figure 5 provides a visual representation of a 250 millisecond delay in real time.

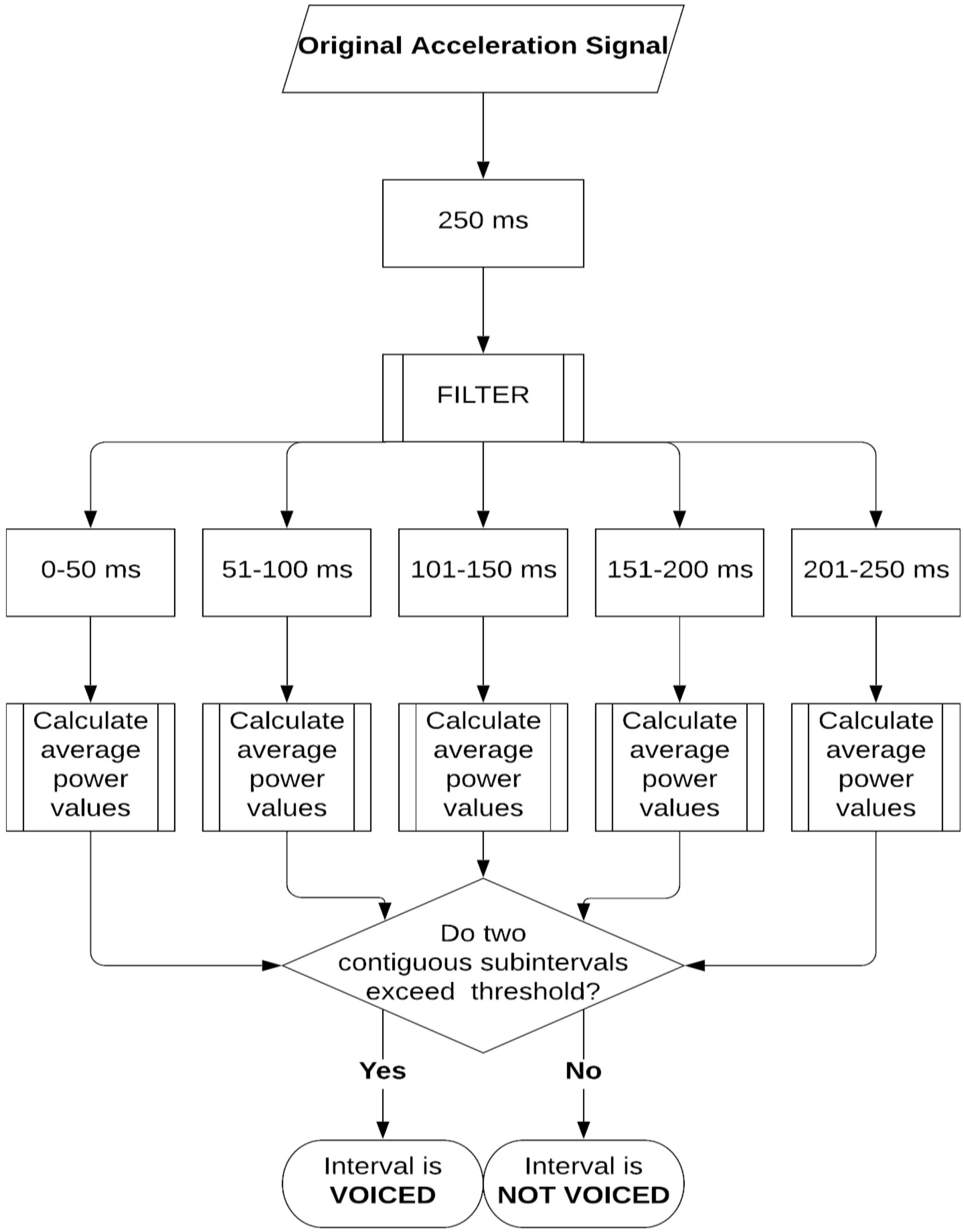

Figure 4. VAD Algorithm: The acceleration signal was analyzed in 250 millisecond intervals, where each interval was divided into 5 subintervals of 50 milliseconds. If two contiguous subintervals had an average power greater than the threshold, the entire interval was considered voiced.

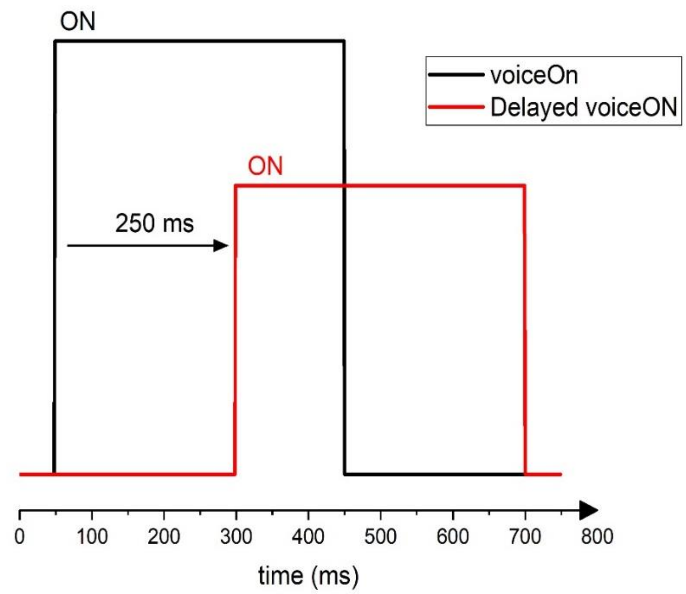

Figure 5. Delay in detection of voice activity onset in real-time based on interval length. The VAD algorithm will analyze each interval completely to determine if it is voiced. In real-time, then, the application of VTS at the onset of speech will be delayed by the length of the interval.

The VAD algorithm output was an integer vector with values greater than zero indicating voiced frames. This vector was plotted against the filtered acceleration signal to visually judge the accuracy of the algorithm (Figure 6). Quantitatively, the number of onsets and total time marked as voiced was calculated. For accuracy validation, the same visual and quantitative process was applied to the audio data and compared to the acceleration values.

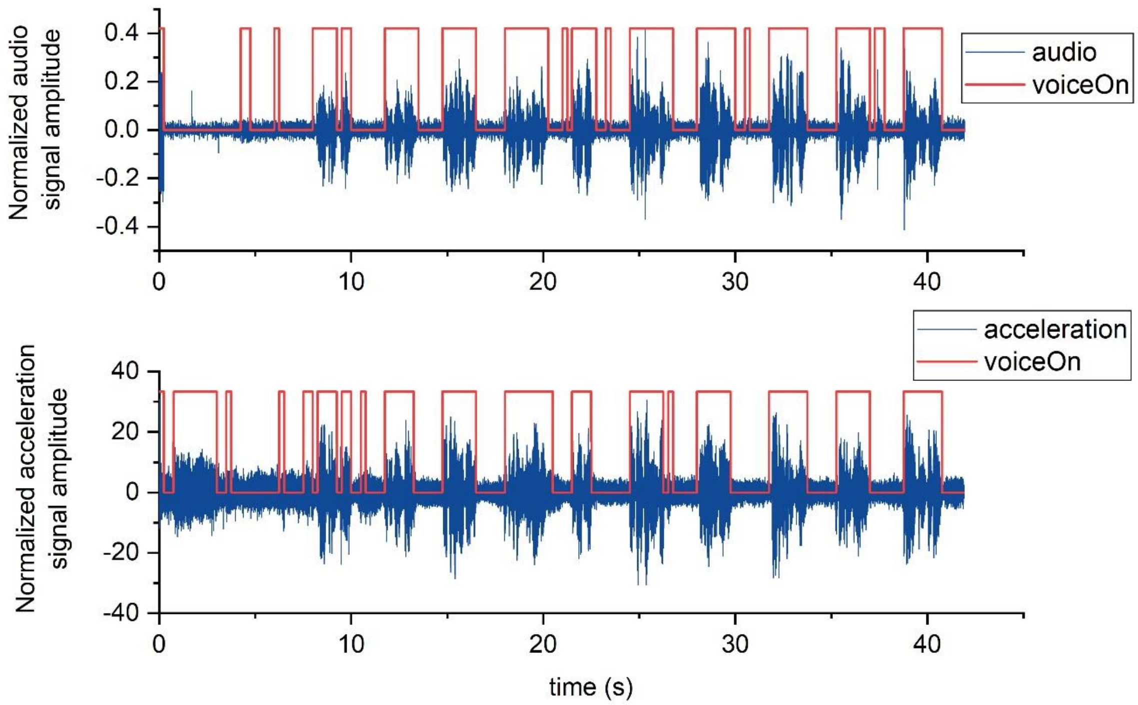

Figure 6. Comparison of audio (above) and acceleration (below) data after application of the VAD algorithm. The orange line at a value above zero represents the algorithm detecting voice. Both data sets are shortened to half the trial length to ease visualization of the width of each voice onset window.

In vibration trials, due to the magnitude shift of the filtered signal, the threshold needed to be scaled up. The exact magnitude distortion factor was undetermined. Multiple scale factors were tested to find one that produced a number of onsets closest to that calculated in the audio data for the same trial. In the VTS collar application, if the algorithm is inaccurate, this would overestimate the duration of speech and maintain delivery of vibration more consistently. Matching the total time voiced measure on the audio trial would underestimate the duration of each instance of speech and provide short, erratic bursts of vibration. This scale factor was calibrated to each trial, but could be calibrated to individuals.

Results

Two main measures were of interest for extraction from the data: the accuracy of the VAD algorithm at the three different locations and its accuracy when comparing non-vibration to vibration trials. Figure 7 shows the mean percent accuracy for number of onsets and time voiced for the different accelerometer locations with and without vibrations. It also shows the distribution of individual data as a comparison of these two variables. The individual data is plotted to show groupings and trends in the data based on presence of vibration and accelerometer location. Percentage accuracy was calculated as the absolute percent error subtracted from 100%. Note that the individual data table (see Appendix) is condensed because of the equality of time and frequency domain calculations and the exclusion of time domain power calculations for vibration trials.

When comparing the different accelerometer attachment locations, the thyroid cartilage has greater than 90% accuracy in both number of onsets and total time voiced, excluding total time voiced in vibration trials (see Figure 7). VAD algorithm accuracy in number of onsets and total time voiced for the sternocleidomastoid shows high variability, ranging from 40% to 74% and 38% to 63% respectively, when including standard error. The C7 accelerometer position recorded below 15% percent accuracy with high variability in both vibration conditions. When comparing non-vibration to vibration trials, the percentage accuracies are similar for number of onsets. The thyroid cartilage and sternocleidomastoid positions decrease in accuracy for total time voiced when adding vibration (see Figure 7). In general, because the number of onsets in the audio data was the metric used to adjust the threshold scale factor for vibration trials, the accuracy for number of onsets is greater in vibration than non-vibration trials and greater than that of calculated time voiced for all vibration trials.

Table 1. Mean percentage accuracy of VAD algorithm for acceleration data compared to audio data with standard error.

	Number of Onsets	Standard Error	Total Time Voiced	Standard Error
No VTS, Thyroid Cartilage	91.67%	2.41%	94.89%	1.58%
No VTS, Sternocleidomastoid	61.42%	12.39%	51.05%	12.39%
No VTS, C7	12.47%	6.70%	5.22%	2.92%
VTS, Thyroid Cartilage	94.88%	2.13%	62.03%	7.96%
VTS, Sternocleidomastoid	73.31%	7.97%	52.17%	12.00%
VTS, C7	12.28%	7.88%	4.66%	2.90%

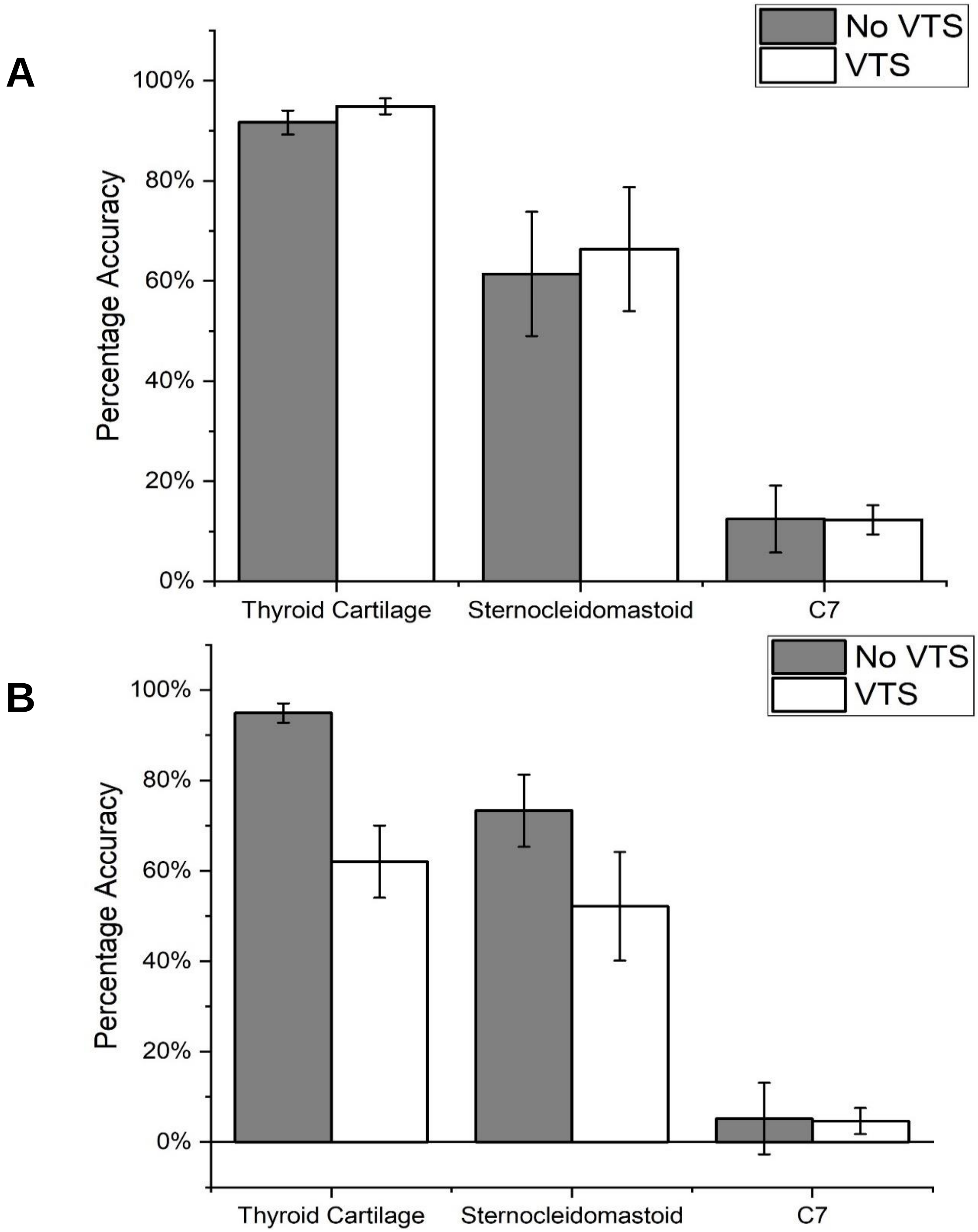

Figure 7. (A) Mean percentage accuracy and standard error for VAD algorithm in calculating Number of Onsets for voice activity at each location for vibration and non-vibration trials. (B) Mean percentage accuracy and standard error for VAD algorithm in calculating Total Time Voiced for voice activity at each location for vibration and non-vibration trials.

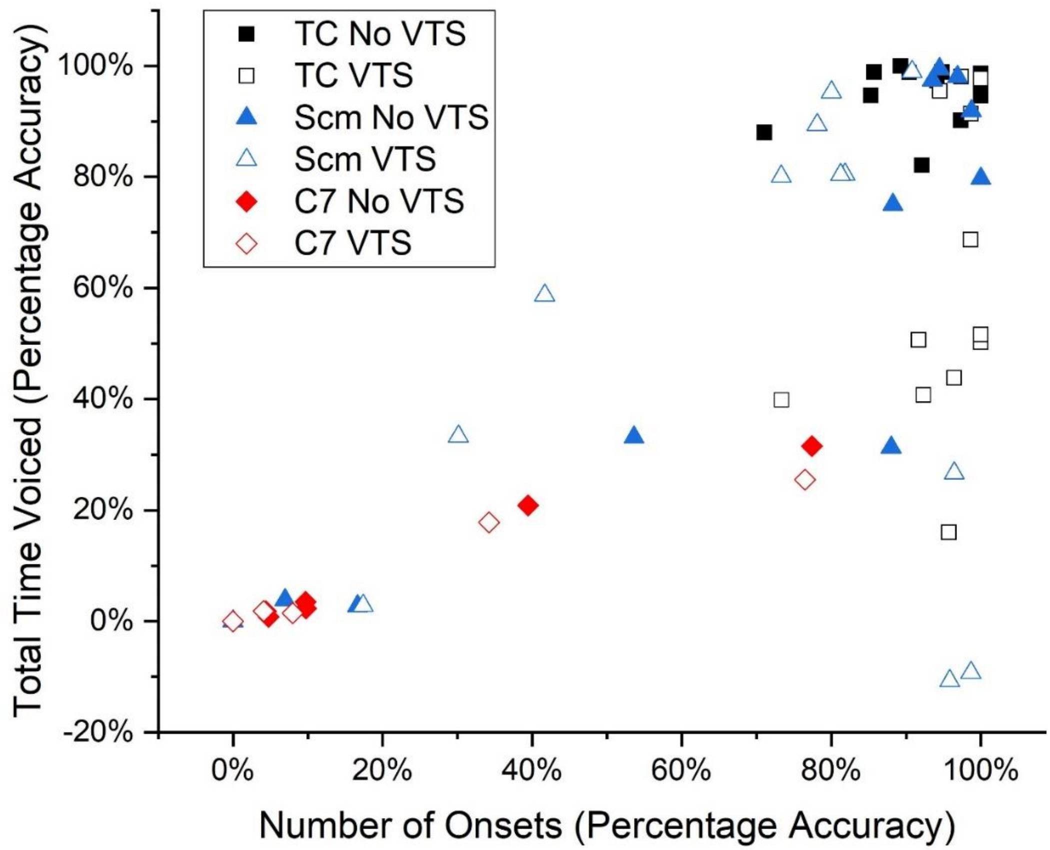

Figure 8. Individual percentage accuracy data plotted as Total Time Voiced against Number of Onsets. Filled shapes represent trials without VTS. Unfilled shapes represent trials with VTS. Negative percentage accuracy occurs when the calculated value, based on acceleration signal, is more than double the actual value, based on audio data. This results in an absolute percent error greater than 100%. Accelerometer attachment locations are abbreviated: TC – Thyroid Cartilage, Scm – sternocleidomastoid, and C7.

The distribution of individual percentage accuracy data as a comparison of total time voiced against number of onsets reflects groupings of data from certain locations (see Figure 8). As expected from the mean data, there is a cluster of high accuracy data for the thyroid cartilage trials without vibration and low accuracy data for nearly all C7 trials. Also evident from the high standard error of the mean data is the large variability in sternocleidomastoid data. The vibration trials at the thyroid cartilage follow a vertical trend. This indicates an accuracy above 70% for number of onsets calculated, but variability in accuracy for total time voiced for the thyroid cartilage location with vibration (see Figure 8).

The individual data table (see Appendix) indicates that the VAD algorithm for the vibration trials often overestimated the total time voiced. However, in trials with less than 33% accuracy, the calculated total time voiced was much less than that of the audio comparison. Trials with significantly less time voiced indicate difficulty in effectively filtering the vibration or alternating current noise, causing a lack of differentiation by the VAD algorithm between voiced and unvoiced signals.

Visualizing the data based on actual acceleration signals presents a clearer understanding of the results VAD algorithm in each condition. The mean percentage data and individual distribution of data shows a decrease in percentage accuracy from the thyroid cartilage as the most accurate to C7 as the least accurate. Figure 9 shows a comparison of acceleration signals from each location in separate trials without vibration.

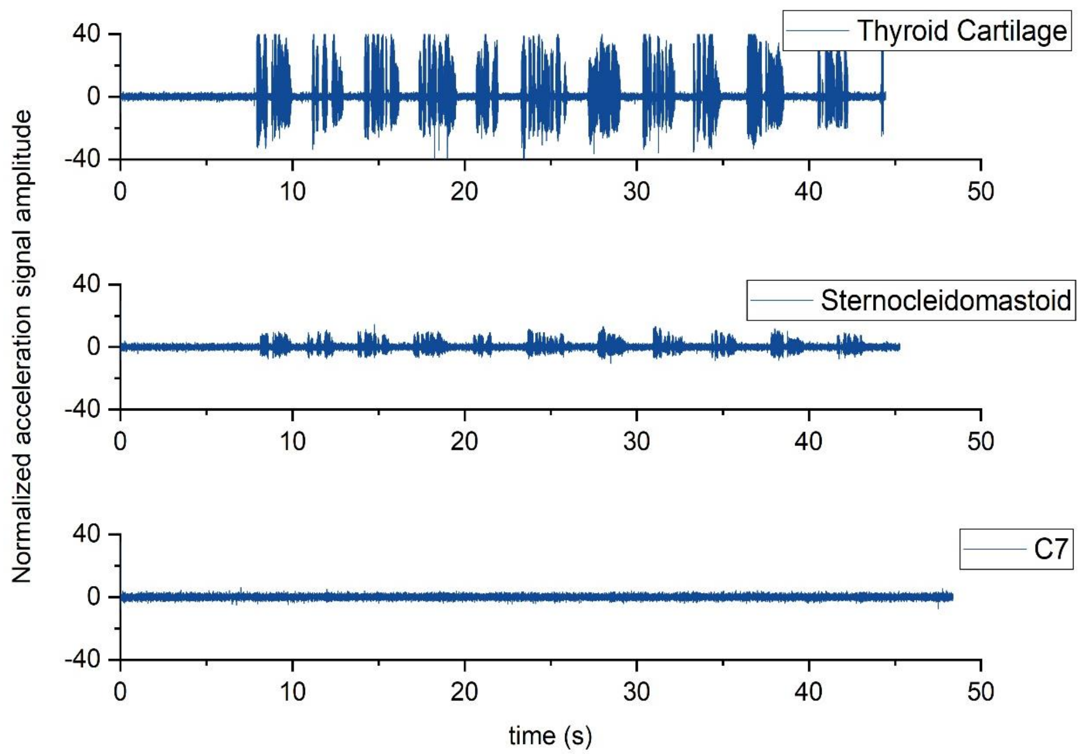

Figure 9. Comparison of acceleration signals, at a normal speed of speech, at each of three accelerometer attachment locations: thyroid cartilage (top), sternocleidomastoid (middle), and C7 (bottom). Each signal is a different trial without vibration. Y-axis scale is intentionally kept constant to emphasize amplitude differences between each signal.

There is a trend of decrease in total acceleration signal amplitude with decrease in VAD algorithm accuracy. There is also a decrease in amplitude of the speech signal in comparison to baseline noise with C7 showing no visually distinguishable voice signal.

Slow and normal speeds of speech were tested to accommodate different speaking styles and disordered voices. The VAD algorithm shows an ability to differentiate sentences and separate words or phrases without overlapping across different instances of speech (see Figure 10).

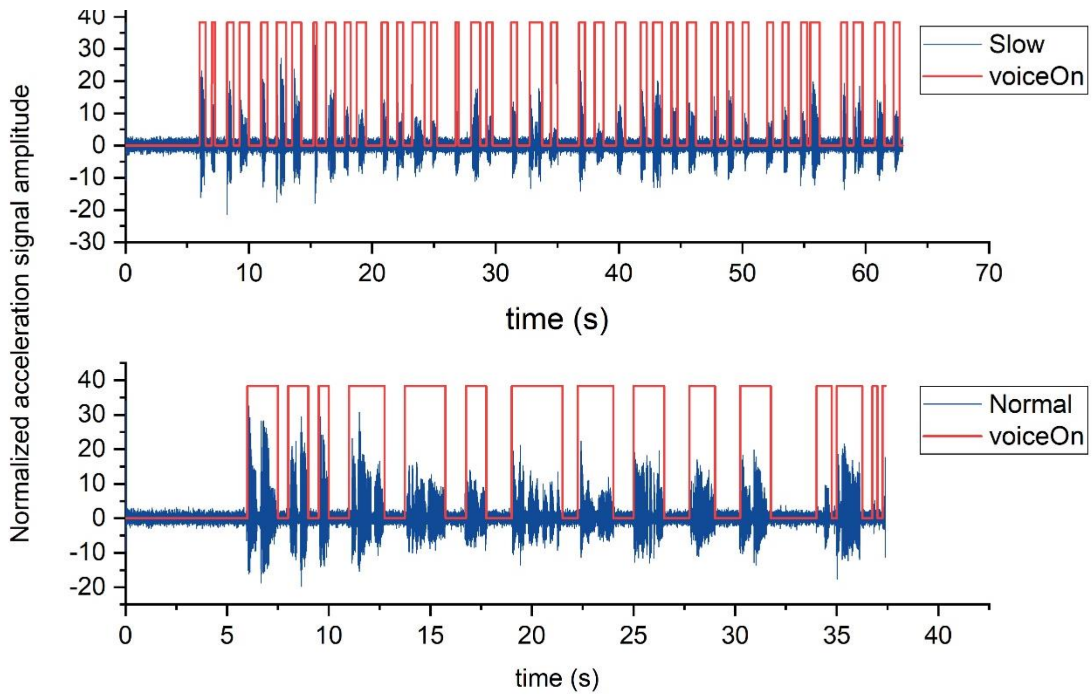

Figure 10. First half of slow (above) and normal (below) speech speed trials plotted with the VAD algorithm output for each trial respectively. In both conditions, the VAD algorithm accurately recognizes the onset and offset of voice activity without overlapping instances of speech, regardless of speed of speech. Each signal is a separate trial without vibration. Half of the trials were plotted for the ease of visualizing the data.

The last condition tested was the application of VTS. Expectedly, introducing VTS to the acceleration recordings added noise to the data. The VAD algorithm was less effective in distinguishing speech activity from silence due to this added noise (see Figure 11).

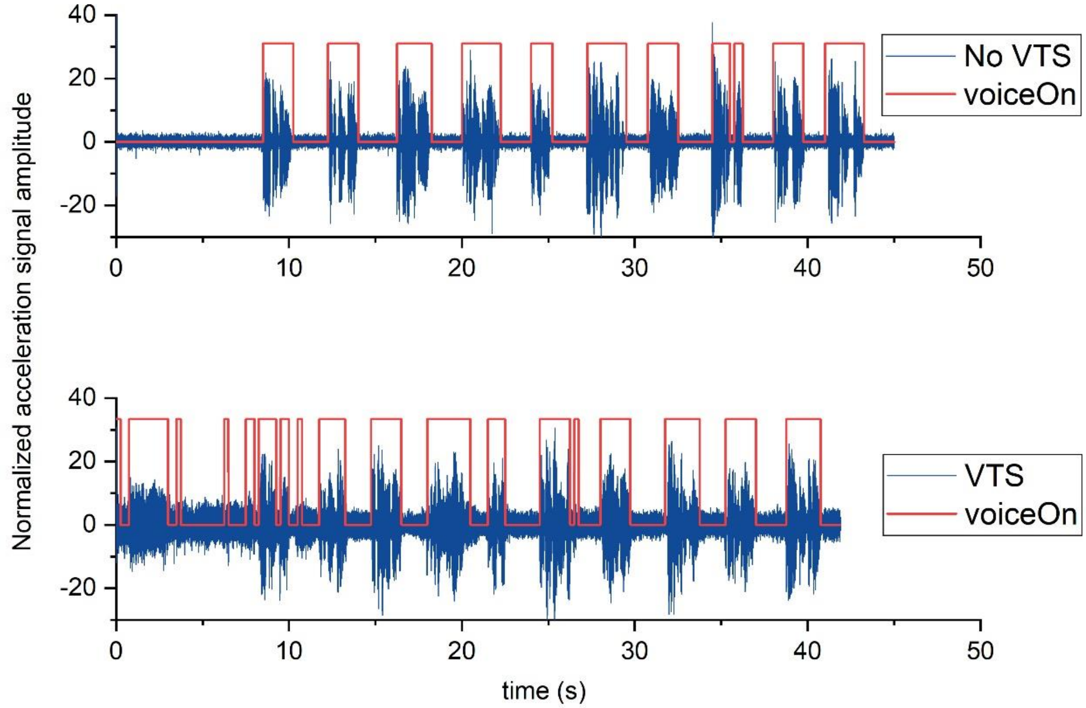

Figure 11. First half of No VTS and VTS trials, at a normal speed of speech, plotted with the VAD algorithm output for each trial respectively. The VAD algorithm is less accurate in detecting the onset and offset of voice due to the increased noise from the vibration trials. The increase in noise due to vibration is evident in the signal as an increased amplitude range during portions without speech compared to the No VTS trial above. Each signal is a separate trial. Half of the trials were plotted for the ease of visualizing the data.

The VTS trial shows a greater amplitude range during portions without speech activity than the No VTS trial (see Figure 11). This increased amplitude range without speech represents the presence of vibration and its associated noise.

Finally, the real-time delay introduced by application of the VAD algorithm can be visualized in the context of recorded acceleration signals. The VAD algorithm plotted with the expected delay in real-time indicates voice activity just after the acceleration signal shows actual voice activity. The VAD algorithm without a delay indicates voice activity just before actual speech (see Figure 12).

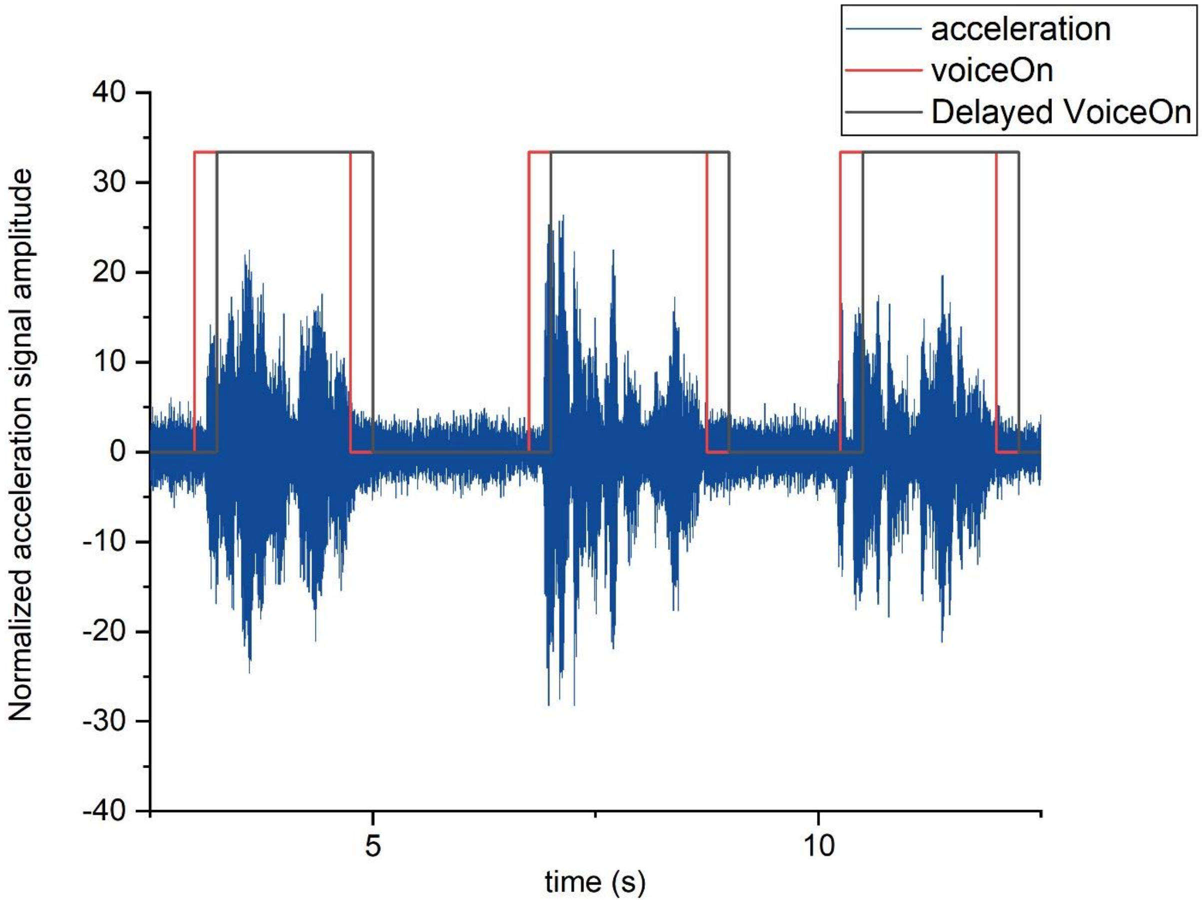

Figure 12. A 15 second interval of acceleration data with the VAD algorithm applied with and without a delay. The delay shown is introduced when applying the VAD algorithm in real-time (see Figure 5). The delayed detection of voice onset indicates voice activity just after the acceleration signal shows actual speech activity.

Discussion

Attachment Location

The first aim of this study was to determine which of three accelerometer attachment locations would provide the best acceleration signal from neck surface vibrations amidst vibration applied to the laryngeal muscles. Based on the results of this study, the thyroid cartilage is the location that provides the acceleration signal with the highest accuracy when compared to the sternocleidomastoid and C7 positions (see Figure 7). The acceleration signals decrease in overall amplitude from the thyroid cartilage to C7 positions. More importantly, the amplitude difference between speech and non-speech signals also decreases (see Figure 9). This can be attributed to the attenuation of the neck surface vibrations caused by speech as the distance from the larynx increases (Moser & Oyer, 1958; Munger & Thomson, 2008). The attenuation of signal amplitude and similarity of speech and non-speech signals causes the VAD algorithm to decrease in accuracy. Comparing individual sternocleidomastoid and C7 data (see Appendix) supports the phenomenon of signal attenuation due to anatomical differences of the neck between participants. Only one participant showed high accuracy, regardless of vibration, at both the sternocleidomastoid and C7 locations. The other participants showed a large decrease in accuracy between the two locations, with the sternocleidomastoid providing a better signal.

The spread of individual data when comparing the calculated number of onsets to total time voiced presents a strong case against use of the sternocleidomastoid and C7 positions for accelerometer-based VAD (see Figure 8). The sternocleidomastoid data was highly variable across participants, regardless of applying VTS. While No VTS trials had greater accuracy in both measures, certain participants failed to produce any distinguishable speech signal at the sternocleidomastoid even without vibration. The C7 attachment location consistently fails to produce an accurately distinguishable speech signal. The individual data from the thyroid cartilage, while variable in total time voiced with VTS applied, still has greater than 70% accuracy in number of onsets for both VTS and No VTS trials (see Figure 8). With the highest accuracy and relatively low variability, compared to the other locations, in both measures, the thyroid cartilage is the best location of the three tested for accelerometer-based VAD.

VAD Algorithm

The second aim of this study was to develop a VAD algorithm with to implement in a device that analyzes neck surface vibrations in real-time to deliver VTS during speech. The VAD algorithm shows high accuracy in the absence of vibration when provided with a strong signal at the thyroid cartilage. Unfortunately, the accuracy of the algorithm in calculating total time voiced at the thyroid cartilage decreases to roughly 60%, similar to the sternocleidomastoid location, when introducing vibration (see Figure 7B). The cause of this decrease in accuracy at the thyroid cartilage is the noise introduced to the acceleration signal by the VTS and the motors producing the vibration. When examining an acceleration signal with VTS applied, the amplitude range of a non-speech segment is greater than that of a similar segment without VTS (see Figure 11). This general increase in amplitude causes the VAD algorithm to mistake non-speech segments for voice activity. These inaccuracies can be seen in Figure 11 around the 10-second mark as short, frequent windows detected as voice activity when there should be a consistent, long window over the course of a spoken sentence.

This decrease in accuracy when adding VTS could be due to the experimental setup and avoidable in specific applications of accelerometer recording. While introducing vibration certainly added noise to the acceleration signal, the frequency of the vibration itself was easily filtered out. The noise due to the alternating current that powered the vibratory motors was more difficult to remove from the signal. The 60 Hz alternating current noise varied slightly as the vibration of the motors was not perfectly consistent. This made filtering the 2^{nd}, 3^{rd}, and 4^{th} harmonics, all of which had large energies relative to the speech signal, difficult. The VAD algorithm may not need to account for this alternating current noise, however, as the real-time application of this system in a collar would be powered by a direct current contained within the device, potentially eliminating the harmonic noise frequencies altogether. If the direct current truly does not have such noise, the VAD algorithm should be robust to the noise caused by introducing vibration and function at an overall higher accuracy than recorded in this study.

Limitations

The aims of this study were accomplished, to a degree, in that the VAD algorithm can easily be implemented into a device for use in real-time with an accelerometer location at the thyroid cartilage. Even so, this study has limitations that have reduced the effectiveness of the VAD algorithm that was developed. An important change to the experimental procedure would have been to record baseline trials with vibration in addition to those without vibration. In this way, the baseline trials with vibration would require the same filters as the speech trials. This methodological change could potentially eliminate the need to scale the threshold due to magnitude distortions caused by filtering the data because both the baseline and speech signals would be filtered. This change could also make it easier to account for the noise caused by the alternating current, since it would be included in the baseline trials as well. Another limitation of the experimental setup was that the microcontroller used to collect data from the accelerometer in this study was not capable of higher frequency recordings. While the fundamental frequency of voice falls within the sampling frequency used in this study, higher frequency recordings could provide speech signals that are more easily isolated from the noise present in signals from this study. A limitation of accelerometers themselves are their inability to capture vibration data from voiced consonants as opposed to vowels, causing a lower intelligibility of the speech signal (Acker-Mills, Houtsma, & Ahroon, 2006). This limitation can account for inaccuracies of the VAD algorithm and suggest a potential path for improvement.

Another limitation of this study considers the specific placements of the accelerometer for each participant. Testing more locations would certainly provide a clearer image of ideal locations to record neck surface vibrations. However, utilizing more distinct landmarks, specifically for the sternocleidomastoid, could yield more consistent results. Munger and Thomson (2008) gathered unique data from multiple locations on the lateral surface of the neck, implying that even small variations in placement of the accelerometer on the sternocleidomastoid could yield noticeably different results. This is supported by the variability in accuracy of the VAD algorithm across participants at the sternocleidomastoid location. Certain participants consistently showed high accuracy, while two presented accuracy below 20% for some trials. Improving the consistency of accelerometer attachment may be an important change in further study. Finally, this experimental setup was limited to using only one accelerometer to record data for each trial. Ideally, one accelerometer would be attached at each location to collect data simultaneously for every trial. In this way, the exact same speech signal could be compared in each location.

Implications for the Wearable VTS Collar

This study was conducted as part of a larger project with the goal of creating a wearable collar that delivers vibration therapy to SD patients as they speak (Mahnan et al., 2019). The findings of this study have important implications for the implementation of VAD in the collar. Regarding the location of the accelerometer attachment in the collar, the thyroid cartilage is the only viable location that was tested. Even with VTS, the thyroid cartilage produced an acceleration signal amplitude from speech that was large enough to be distinguished from non-voice activity. The sternocleidomastoid was far too variable across participants to be feasible in a collar intended for a large population. C7 could be expected not to produce a usable signal for VAD in most SD patients. With a refined VAD algorithm, neck-surface vibrations from the thyroid cartilage should be effective in providing an accurate representation of voice activity.

The results of the VAD algorithm used in this study provide valuable insight for its use in the actual device. The modular aspects of the algorithm were the length of intervals and the scale factor used to account for magnitude distortion of acceleration signals after filtering. While modular aspects of the algorithm allow for flexibility in adjusting the accuracy of VAD, they require individual calibration to set at accurate values. Fortunately, both of these modular aspects could potentially be unnecessary in the collar. The need for a scale factor could be removed if advanced signal processing, such as an adaptive filter, was implemented. The 250 millisecond interval length used in this study could also be maintained in the collar. The VAD algorithm effectively detects voice activity when full sentences and single words or short phrases are spoken (see Figure 10). This is evidence that the interval length is the correct length to accommodate normal speech and slow speech with the appropriate delivery of VTS. A longer interval length could cause overlap between different instances of speech while a shorter interval could cause more erratic voice activity detection. The interval length is also short enough to prevent introducing a delay in real-time that is too large. A delay that is too large would cause vibration to begin noticeably later than the actual onset of speech. This could be uncomfortable and ineffective in improving speech quality for the collar wearer. The 250 millisecond delay, however, causes voice activity onset detection just after the actual initiation of speech activity, as seen by the acceleration signal (see Figure 12). Based on these results, the VAD algorithm, if improved in signal processing, could easily be implemented in real-time in the wearable VTS collar.

Future Directions

In light of the implications for the wearable VTS collar and limitations of the study, there are many future directions to expand this study. Considering the target population of the VTS therapy device, there should be further testing of this VAD algorithm and the feasibility of using an accelerometer to capture speech at varying volumes. Beyond this, the algorithm should be tested on data collected from patients of SD. While Cheyne et al. (2003) showed that an accelerometer was capable of capturing acceleration data from speech in dysphonic patients, real-time voice activity detection remains undetected in the SD patient population. The concept of VAD using an accelerometer could even be explored outside of this specific application of delivering vibration therapy to SD patients. Mehta et al. (2013) has shown the potential for using long-term tracking of neck surface vibration data for the diagnosis of speech disorders.

The algorithm itself is in need of continued development. The current method for filtering out noise from vibration uses multiple filters, which could require more computational power than available in the wearable collar device. The filtering is also decreases in accuracy with the application of VTS. There are a multitude of ways to design and apply various filters as well as signal processing metrics that could be used to solve both issues. An adaptive filter, for example, is a more advanced method of treating the acceleration data that could eliminate the need for individual calibration. Rather, it would adjust to each participant based on how it was designed. Also, an adaptive filter would likely filter out noise from VTS better than the filtering used in this study, resulting in improved accuracy of the VAD algorithm. New methods for analyzing signal properties are still being developed that could be used in place of the average power calculations in this study. Ultimately, while there are numerous possible methods for improving the VAD algorithm that was developed, it was outside the scope of this study to find the best one.

Conclusion

In summary, this study examined the feasibility of recording neck surface vibrations with an accelerometer at various locations to develop a VAD algorithm for use in a collar to deliver VTS therapy to SD patients. This study showed that the thyroid cartilage was the best location for recording neck surface vibrations, but that there is also potential at certain regions along the sternocleidomastoid muscle. The VAD algorithm functions with high accuracy at the thyroid cartilage without vibration, but shows a decrease in accuracy regarding calculated total voiced time upon the introduction of vibration.

Fortunately, there is the potential for the accuracy of this algorithm to increase if the direct current in the device application does not produce the same noise as the alternating current in the experimental setup.

Regarding the application of these results to the actual collar, the thyroid cartilage attachment location is the most feasible location for accelerometer-based VAD of the tested locations. The VAD algorithm, if its signal processing workflow is improved, could be implemented as is in real-time. It would effectively detect the onset and offset of voice activity using neck-surface vibrations from the thyroid cartilage.

Further research should evaluate variations in voice such as changes in volume and patients of voice disorders. These new variables will require improvement of the VAD algorithm to accommodate a wider variety of speech signals. To do so, more advanced methods of signal filtering as well as signal analysis should be pursued.

Bibliography

Acker-Mills, B. E., Houtsma, A. J., & Ahroon, W. A. (2006). Speech intelligibility in noise using throat and acoustic microphones. *Aviation, space, and environmental medicine, 77*(1), 26-31.

Askenfelt, A., Gauffin, J., Sundberg, J., & Kitzing, P. (1980). A comparison of contact microphone and electroglottograph for the measurement of vocal fundamental frequency. *Journal of Speech, Language, and Hearing Research, 23*(2), 258-273.

Audacity Team. (2019). Audacity(R): Free Audio Editor and Recorder (Version 2.3.2).

Cheyne, H. A., Hanson, H. M., Genereux, R. P., Stevens, K. N., & Hillman, R. E. (2003). Development and testing of a portable vocal accumulator. *Journal of Speech, Language, and Hearing Research, 46*(6), 1457-1467.

Hillman, R. E., Heaton, J. T., Masaki, A., Zeitels, S. M., & Cheyne, H. A. (2006). Ambulatory monitoring of disordered voices. *Annals of Otology, Rhinology & Laryngology, 115*(11), 795-801.

Khosravani, S., Mahnan, A., Yeh, I.-L., Aman, J. E., Watson, P. J., Zhang, Y., . . . Konczak, J. (2019). Laryngeal vibration as a non-invasive neuromodulation therapy to improve voice symptoms in spasmodic dysphonia. *Submitted*.

Lindstrom, F., Ren, K., Li, H., & Waye, K. P. (2009). Comparison of two methods of voice activity detection in field studies. *Journal of Speech, Language, and Hearing Research*.

Mahnan, A., Faraji, S. A., & Konczak, J. (2019). *Wearable non-invasive neuromodulation device for the symptomatic treatment of the voice disorder spasmodic dysphonia*. Paper presented at the Design of medical devices, Minneapolis, MN.

Matic, A., Osmani, V., & Mayora, O. (2012). *Speech activity detection using accelerometer.* Paper presented at the Engineering in medicine and biology society (EMBC), 2012 annual international conference of the IEEE.

Mehta, D. D., Zanartu, M., Feng, S. W., Cheyne II, H. A., & Hillman, R. E. (2012). Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform. *IEEE Transactions on Biomedical Engineering, 59*(11), 3090-3096.

Mehta, D. D., Zañartu, M., Van Stan, J. H., Feng, S. W., Cheyne, H. A., & Hillman, R. E. (2013). *Smartphone-based detection of voice disorders by long-term monitoring of neck acceleration features.* Paper presented at the 2013 IEEE International Conference on Body Sensor Networks.

Moser, H. M., & Oyer, H. J. (1958). Relative intensities of sounds at various anatomical locations of the head and neck during phonation of the vowels. *The Journal of the Acoustical Society of America, 30*(4), 275-277.

Munger, J. B., & Thomson, S. L. (2008). Frequency response of the skin on the head and neck during production of selected speech sounds. *The Journal of the Acoustical Society of America, 124*(6), 4001-4012.

Nacci, A., Fattori, B., Mancini, V., Panicucci, E., Ursino, F., Cartaino, F., & Berrettini, S. (2013). The use and role of the Ambulatory Phonation Monitor (APM) in voice assessment. *ACTA otorhinolaryngologica italica, 33*(1), 49.

Nolan, M., Madden, B., & Burke, E. (2009). *Accelerometer based measurement for the mapping of neck surface vibrations during vocalized speech.* Paper presented at the 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society.

Schafer, R. W., & Oppenheim, A. V. (2010). *Discrete-time signal processing* (3 ed.). Upper Saddle River, NJ: Pearson.

Stevens, K. N. (2000). Acoustic phonetics (Vol. 30, pp. 261-266): MIT press.

Švec, J. G., Titze, I. R., & Popolo, P. S. (2005). Estimation of sound pressure levels of voiced speech from skin vibration of the neck. *The Journal of the Acoustical Society of America, 117*(3), 1386-1394.

Szabo, A., Hammarberg, B., Hakansson, A., & Sodersten, M. (2001). A voice accumulator device: Evaluation based on studio and field recordings. *Logopedics Phoniatrics Vocology, 26*(3), 102-117.

Vitikainen, A. M., Makela, E., Lioumis, P., Jousmaki, V., & Makela, J. P. (2015). Accelerometer-based automatic voice onset detection in speech mapping with navigated repetitive transcranial magnetic stimulation. *J Neurosci Methods, 253*, 70-77. doi:10.1016/j.jneumeth.2015.05.015

Yiu, E. M., & Yip, P. P. (2016). Effect of noise on vocal loudness and pitch in natural environments: an accelerometer (ambulatory phonation monitor) study. *Journal of Voice, 30*(4), 389-393.

Appendix

Table 2. Individual participant data for each trial. Trial names are abbreviated: N – Normal speed of speech, S – slow speed of speech, TC – Thyroid cartilage, Scm – sternocleidomastoid, and C7. Percentage accuracy was calculated as the absolute percent error subtracted from 100%. Negative percentage accuracy values indicate that the percentage error was greater than 100%, meaning the recorded value was more than double the actual value.

		Acceleration - Average Spectral Power			Audio - Average Spectral Power		Percentage Accuracy	
Participant	**Trial**	**Window Count**	**Time Voiced (s)**	**Scale Factor**	**Window Count**	**Time On (s)**	**Number of Onsets**	**Total Time Voiced**
	noVTS_N_TC	29	49.75	1	34	47.25	85.29%	94.71%
	VTS_N_TC	28	73.25	35	26	46	92.31%	40.76%
	noVTS_S_TC	70	56	1	76	47.5	92.11%	82.11%
	VTS_S_TC	77	50.5	35	75	51.5	97.33%	98.06%
	noVTS_N_Scm	34	44.25	1	36	44	94.44%	99.43%
	VTS_N_Scm	29	11	15	28	41.25	96.43%	26.67%
	noVTS_S_Scm	72	49.5	1	77	48.25	93.51%	97.41%
	VTS_S_Scm	69	46.25	8	76	46.75	90.79%	98.93%
	noVTS_N_C7	4	1	1	41	42.75	9.76%	2.34%
S05	noVTS_S_C7	7	1.75	1	72	49.75	9.72%	3.52%
	noVTS_N_TC	30	45.75	1	35	45.25	85.71%	98.90%
	VTS_N_TC	38	69.25	4	30	43.25	73.33%	39.88%
	noVTS_S_TC	76	52.25	1	76	50.5	100.00%	96.53%
	VTS_S_TC	73	76.75	9	73	51.25	100.00%	50.24%
	noVTS_N_Scm	31	37.75	1	32	38.5	96.88%	98.05%
	VTS_N_Scm	23	88.5	1	24	42	95.83%	-10.71%
S07	noVTS_S_Scm	76	42.5	1	77	46.25	98.70%	91.89%

Table 2 - Continued

	VTS_S_Scm	76	107.25	2	75	51.25	98.67%	-9.27%
	noVTS_N_C7	0	0	1	29	39.5	0.00%	0.00%
	VTS_N_C7	0	0	1	27	39	0.00%	0.00%
	noVTS_S_C7	0	0	1	74	48.75	0.00%	0.00%
	VTS_S_C7	0	0	1	74	52.25	0.00%	0.00%
	noVTS_N_TC	32	38.75	1	34	37.75	94.12%	97.35%
	VTS_N_TC	27	59.75	15	28	38.25	96.43%	43.79%
	noVTS_S_TC	76	44.75	1	74	40.75	97.30%	90.18%
	VTS_S_TC	72	53.5	18	73	40.75	98.63%	68.71%
	noVTS_N_Scm	22	10.25	1	25	32.75	88.00%	31.30%
	VTS_N_Scm	26	24.75	1	22	30.75	81.82%	80.49%
	noVTS_S_Scm	37	12.75	1	69	38.5	53.62%	33.12%
	VTS_S_Scm	52	40	1	65	42	80.00%	95.24%
	noVTS_N_C7	1	0.25	1	21	32	4.76%	0.78%
	VTS_N_C7	0	0	1	28	29.75	0.00%	0.00%
	noVTS_S_C7	3	0.75	1	68	41.5	4.41%	1.81%
S08	VTS_S_C7	0	0	1	72	40.25	0.00%	0.00%
	noVTS_N_TC	24	36	1	24	36.5	100.00%	98.63%
	VTS_N_TC	22	56.75	10	22	38.25	100.00%	51.63%
	noVTS_S_TC	66	42.25	1	73	42.75	90.41%	98.83%
	VTS_S_TC	73	83.25	51	70	45.25	95.71%	16.02%
	noVTS_N_Scm	0	0	1	28	35.75	0.00%	0.00%
	VTS_N_Scm	4	1	4	23	35	17.39%	2.86%
S09	noVTS_S_Scm	0	0	1	74	40.5	0.00%	0.00%

Table 2 - Continued

	VTS_S_Scm	22	13.5	3	73	40.5	30.14%	33.33%
	noVTS_N_C7	0	0	1	26	33.75	0.00%	0.00%
	VTS_N_C7	0	0	1	26	32.5	0.00%	0.00%
	noVTS_S_C7	0	0	1	71	39	0.00%	0.00%
	VTS_S_C7	0	0	1	76	40	0.00%	0.00%
	noVTS_N_TC	27	38.5	1	38	43.75	71.05%	88.00%
	VTS_N_TC	35	42.5	4	35	43.5	100.00%	97.70%
	noVTS_S_TC	73	45.25	1	77	44.75	94.81%	98.88%
	VTS_S_TC	76	42.5	6	75	46.5	98.67%	91.40%
	noVTS_N_Scm	38	26.25	1	34	35	88.24%	75.00%
	VTS_N_Scm	26	44.25	3	32	37	81.25%	80.41%
	noVTS_S_Scm	73	32.5	1	73	40.75	100.00%	79.75%
	VTS_S_Scm	3	52.75	3	73	44	73.31%	80.11%
	noVTS_N_C7	38	11.75	1	31	37.25	77.42%	31.54%
	VTS_N_C7	26	9.25	1	34	36.25	76.47%	25.52%
	noVTS_S_C7	30	8.5	1	76	40.75	39.47%	20.86%
S10	VTS_S_C7	25	7.25	1	73	40.75	34.25%	17.79%
	noVTS_N_TC	25	38.25	1	28	38.25	89.29%	100.00%
	VTS_N_TC	26	54.5	13	24	36.5	91.67%	50.68%
	noVTS_S_TC	73	44	1	73	41.75	100.00%	94.61%
	VTS_S_TC	69	46.25	20	73	44.25	94.52%	95.48%
	noVTS_N_Scm	4	1	1	24	36.25	16.67%	2.76%
	VTS_N_Scm	10	20.25	1	24	34.5	41.67%	58.70%
S11	noVTS_S_Scm	5	1.5	1	72	39	6.94%	3.85%

Table 2 - Continued

	VTS_S_Scm	57	37.75	1	73	42.25	78.08%	89.35%
	noVTS_N_C7	0	0	1	29	32.25	0.00%	0.00%
	VTS_N_C7	2	0.5	1	25	34	8.00%	1.47%
	noVTS_S_C7	3	0.75	1	73	40.75	4.11%	1.84%
	VTS_S_C7	3	0.75	1	73	41.5	4.11%	1.81%

SUPPLEMENTARY MATERIALS

Subject Information Form

University of Minnesota – Human Sensorimotor Control Laboratory

Human Sensorimotor Control Laboratory

Research Participant Information Form

GENERAL INFORMATION

Study Name: ____

Subject Category (e.g., control, patient): ____________________________________

Subject Name:_______________________________________ (first, last)

Subject Gender: Female / Male

Assigned Subject Number:_____________________________

Testing Date:_______________

Birthdate:______________ (dd-mm-yyyy) Age:____________ (years, months)

Handedness Score (attach questionnaire):_------------_

Any other information?

MEDICATION

Are you currently taking any medication? Y N

If so, list medication: ____

MEDICAL HISTORY

Do you have any history of:

Diabetes: Y N If so, since when: ________

Central Nervous System Disease: Y N If so, list approximate dates: ________

Describe:

Peripheral Nerve Disease: Y N If so, list approximate dates: ________

Describe:

Arm Fractures/Luxations: Y N If so, list approximate dates: ________

Right Upper Limb Pathologies (describe):

Describe:

Left Upper Limb Pathologies (describe):

Describe:

Have you had any voice related disorders? Y N If so, describe:

Subject accepted for study (investigator's initials):_________

Test Sentences

I. Speech Tasks – Normal Speed
Take an average breath and read each sentence in what you consider to be your normal conversational speaking style.

1. Tom wants to be in the army.
2. We eat eels every day.
3. He was angry about it all year.
4. I hurt my arm on the iron bar.
5. Are the olives large?
6. John argued ardently about honesty.
7. We mow our lawn all year.
8. Jane got an apple for Ollie.
9. A dog dug a new bone.
10. Everyone wants to be in the army.
11. He is hiding behind the house.
12. Patty helped Kathy carve the turkey.
13. Harry is happy because he has a new horse.
14. During babyhood he had only half a head of hair.
15. Who says a mahogany highboy isn't heavy?
16. Boys were singing songs outside of our house.
17. The puppy hit the tape.
18. See, there's a horse across the street.
19. Sally fell asleep in the soft chair.
20. The policy was suggested in an essay on peace.

II. Speech Tasks – Slow Speed
Take an average breath and read each sentence in a slower, choppy style of speaking, pausing between words or phrases at each "-"

1. Tom – wants – to be – in the army.
2. We – eat eels – every day.
3. He – was angry – about it – all year.
4. I hurt – my arm – on the iron – bar.
5. Are – the olives – large?
6. John – argued ardently – about honesty.
7. We mow – our lawn – all year.
8. Jane – got an apple – for Ollie.
9. A dog – dug – a new – bone.
10. Everyone – wants – to be – in the army.
11. He – is hiding – behind – the house.
12. Patty – helped – Kathy carve – the turkey.
13. Harry – is happy – because he has – a new horse.
14. During babyhood – he had – only half – a head of hair.
15. Who says – a mahogany highboy – isn't heavy?

16. Boys – were singing songs – outside – of our house.
17. The puppy – hit – the tape.
18. See, – there’s a horse – across – the street.
19. Sally – fell asleep – in the soft – chair.
20. The policy – was suggested – in an essay – on peace